적응적 특징추출을 이용한 Radial Basis Function 신경망의 성능개선

조 용 현[†]

요 약

본 논문에서는 적응적으로 추출된 입력 데이터의 특징을 은닉층 뉴런 개수와 중심값 설정에 이용하는 새로운 radial basis 함수 신경망을 제안하였다. 제안된 신경망에서는 입력 데이터의 특징을 효과적으로 추출하기 위해 적응 학습알고리즘의 주요성분분석 기법을 이용하였다. 이렇게 하면 주요성분분석 기법이 가지는 대용량의 입력데이터를 통계적으로 독립인 특징들의 집합으로 변환시키는 장점과 RBF 신경망이 가지는 우수한 속성을 그대로 살릴 수 있다. 제안된 기법의 radial basis 함수 신경망을 200명의 암환자를 2부류(초기와 악성)로 분류하는 문제에 적용하여 시뮬레이션한 결과, k-평균 군집화 알고리즘을 이용한 radial basis 함수 신경망에 의한 결과와 비교할 때 학습시간과 시험 데이터의 분류에서 더욱 우수한 성능이 있음을 확인할 수 있었다. 그리고 신경망의 초기 연결가중치에 대한 의존도와 평활요소의 설정여유도 측면에서도 우수한 특성이 있음을 확인할 수 있었다.

Performance Improvement of Radial Basis Function Neural Networks Using Adaptive Feature Extraction

Yong-Hyun Cho[†]

ABSTRACT

This paper proposes a new RBF neural network that determines the number and the center of hidden neurons based on the adaptive feature extraction for the input data. The principal component analysis is applied for extracting adaptively the features by reducing the dimension of the given input data. It can simultaneously achieve a superior property of both the principal component analysis by mapping input data into set of statistically independent features and the RBF neural networks. The proposed neural networks has been applied to classify the 200 breast cancer databases by 2-class. The simulation results shows that the proposed neural networks has better performances of the learning time and the classification for test data, in comparison with those using the k-means clustering algorithm. And it is affected less than the k-means clustering algorithm by the initial weight setting and the scope of the smoothing factor.

1. 서 론

컴퓨터비젼이나 로보틱스 등과 같은 여러 응용분야에 이용되고 있는 순차적 기법의 대안으로 최근 신경망이 널리 활용되고 있다. 이러한 신경망은 단순한 처리능력을 가지는 뉴런(neuron)들이 대규모로 상호 연결되어 병렬분산처리, 일반화 그리고 학습 등의 속성을 가지고 있어 조합형 최적화 문제나 인식 등과 같은 문제들을 해결하는 데에 많이 이용되고 있다[1-9].

일반적으로 기존의 순차적 기법에서는 입력되는 데이터들에 대한 확률밀도함수(probability density function) 등이 사전에 요구되거나 또는 모든 데이터들의 분포가 정규밀도를 가져야 한다는 제약을 가지고 있다. 특히 문제의 규모가 증가할수록 계산시간이 지나치게 증가되거나 사전에 요구되는 확률밀도함

† 정회원, 대구효성가톨릭대학교 공과대학 컴퓨터정보통신공학부

수 등을 이용할 수 없는 문제점을 내포하고 있다.

이러한 기존 기법들이 가지는 문제점을 해결하기 위해서 입력과 출력간의 사상(mapping)을 가능하게 하는 신경망이 널리 연구되고 있다. 신경망을 이용하는 기법은 대규모의 병렬계산과 분산된 국부적 계산의 속성 면에서는 통계적인 기법이나 계산 이론적인 방법의 문제들도 해결할 수 있다. 그러나 학습패턴과 학습알고리즘이 주어진다고 모든 응용분야를 다 학습시킬 수는 없으며, 일반적으로 신경망은 망의 구조, 뉴런의 활성화함수, 그리고 학습규칙 등에 따라 여러 가지로 나누어진다. 지금까지 널리 이용되고 있는 신경망[2-4]으로는 홉필드 모델(Hopfield model), Kohonen의 자기조직화 특징지도(self-organizing feature map : SOFM), 적응공명이론(adaptive resonance theory : ART) 모델, 그리고 다층전향신경망(layered feedforward neural network) 등이 있다. 이들 중에서 홉필드 모델은 이진입력을 가지는 지도학습(supervised learning)의 신경망이며, SOFM과 ART 모델은 연속입력을 가지는 자율학습(unsupervised learning)의 신경망이다. 또한 다층전향신경망은 연속입력을 대상으로 하는 지도학습의 신경망이다.

한편 입력층과 출력층 사이에 은닉층을 가지는 다층전향신경망은 입출력 데이터 내에 포함된 어떤 비선형 연속함수를 근사화하거나 재구성할 수 있어 대단히 일반적이면서도 융통성 있는 속성을 가지고 있다. 이러한 다층전향신경망으로는 역전파(backpropagation : BP) 알고리즘의 다층신경망(multilayer perceptron : MLP)과 radial basis function(RBF)을 이용하는 신경망 등이 주로 이용된다[2,3]. 이들 중 학습 알고리즘으로 역전파 알고리즘을 이용하는 다층신경망이 지금까지 가장 널리 이용되어 왔다[5]. 이는 MLP가 충분한 뉴런을 가지고 있을 때 어떤 임의의 함수를 근사화시킬 수 있다고 알려져 있으며, 지금까지 많은 시뮬레이터가 개발되어 있기 때문이다. 하지만 MLP는 프로세서의 유용한 모델을 얻기 위해서 많은 학습데이터와 시험데이터를 요구하며, 비파라미터 모델(non-parameter model)이라는 단점을 가지고 있다. 특히, 기울기하강(steepest descent)의 속성을 이용하는 역전파 알고리즘은 학습파라미터의 설정에 따라 수렴속도와 견실성 중 하나 이상의 문제를 가지며, 전역최소점으로의 수렴이 보장되어 있지 않다. 이러한 문제를 해결하기 위한 여러 방법들이 많은 연구자들에 의해서 연구되어 왔다[2-4]. 그 중에서 인간두뇌의 생물학적 뉴런의 기능을 모방한 국부적으로 동조된(locally-tuned) 처리장치를 은닉층의 뉴런으로 이용하는 RBF 신경망이 최근 들어 많은 각광을 받고 있다[7-9]. 이는 RBF 신경망도 MLP와 같이 비선형의 다층전향신경망일 뿐만 아니라 일반적인 근사기(universal approximator)이기 때문이다. 특히 RBF 신경망은 MLP에 비해 훨씬 빠른 수렴속성과 전역최소로의 수렴이 보장된다고 알려져 있다. 그러나 RBF 신경망에서도 은닉층 뉴런의 개수와 함수의 중심(center), 평활요소(smooth factor), 그리고 연결가중치(synaptic weight) 등의 파라미터들이 설정되어야 한다. 이들 파라미터 중에서 학습 중에 은닉층 뉴런의 개수와 함수의 중심 및 평활요소는 고정된 값이며 연결가중치 만이 가변되는 파라미터이다. 특히, 은닉층 뉴런의 개수와 함수의 중심값 결정은 신경망의 구조를 결정하며, 또한 신경망의 성능에도 많은 영향을 미친다. 이를 결정하기 위한 여러 가지 기법들이 제안되고 있으나 일반적으로 입력패턴의 차원이 증가함에 따라 은닉층 뉴런의 수와 중심을 결정하기 위한 계산량은 기하급수적으로 증가한다. 따라서 그 계산량을 줄이기 위한 노력들로 랜덤(random) 선택법, k-평균 군집화(k-means clustering) 알고리즘, 그리고 Kohonen의 자기조직화 특징지도 등이 이용되고 있다[9]. 그러나 대부분의 이들 기법들은 자율적인 방법으로 은닉층의 파라미터들을 결정한 다음 지도학습의 방법으로 출력 연결가중치를 결정한다. 여기서 자율적인 방법은 실제 신경망을 응용할 때 사용자나 군집화 알고리즘 둘 중 어느 하나에 의해서 조정되어야 될 파라미터들이 너무 많다는 단점을 가지고 있으며, 이와 함께 문제내의 국부적인 복잡성도 잘 반영하지 못하는 단점도 가지고 있다. 따라서 근본적으로 입력 데이터가 가지는 차원을 감소시킬 수 있는 새로운 기법을 도입함으로써 결정해야 될 은닉층의 파라미터들을 줄일 수 있을 것이다. 즉, 주어진 입력 데이터의 차원을 감소시켜 그에 해당되는 특징값을 RBF 신경망의 은닉층 파라미터로 사용함으로써 이들을 결정하는데 요구되는 계산량을 효과적으로 감소시킬 수 있을 것이다.

본 연구에서는 수치적인 데이터 집합의 차원을 해석하여 이를 감소시키는 것으로 널리 알려진 주요성분분석(principal component analysis : PCA) 기법

[4-7]을 이용하여 특징값을 추출한 다음, 추출된 특징값을 RBF 신경망의 은닉층 뉴런의 수와 중심으로 이용하는 새로운 조합형 기법을 제안한다. 결국 이렇게 하면 PCA가 가지는 대용량의 입력데이터를 통계적으로 독립인 특징들의 집합으로 변환시키는 장점과 RBF가 가지는 우수한 속성을 그대로 살릴 수 있다. 제안된 기법의 RBF 신경망을 실제 699명의 암환자[10] 중 200명의 환자를 2부류(초기와 악성)로 분류하는 문제에 적용하여 시뮬레이션하고 그 타당성을 확인하였으며, 지금까지 널리 이용되고 있는 k-평균 군집화 알고리즘을 각각 RBF 신경망에 적용한 결과와 비교 고찰하였다.

2. 주요성분분석 기법을 이용한 radial basis 함수의 중심설정

2.1 RBF 신경망

인식이나 분류 등의 문제들을 해결하기 위한 MLP의 대안으로 radial basis 함수를 이용한 RBF 신경망이 널리 이용되고 있다[2,3,4]. MLP와 RBF 신경망 사이에는 유사한 특성이 있는 반면에 은닉층의 수와 같은 구조나 각 은닉층 및 출력층 뉴런의 계산방법과 입·출력 사상에 대한 근사화 등은 아주 다른 속성을 가지고 있다. 그림 1은 3층 전향 신경망인 MLP와 RBF 신경망을 나타낸 것이다. 일반적으로 그림 (a)의 MLP 신경망에서 은닉층 뉴런과 출력층 뉴런의 활성화 함수는 단조함수이며, 그림 (b)의 은닉층 뉴런은 radial basis 함수이고 출력층 뉴런은 단순히 가중된 합만을 출력한다. 여기서 radial basis 함수는 입력벡터 **x**와 중심 **c** 사이의 거리 $r = \| \mathbf{x} - \mathbf{c} \|$ 에 의존하는 다차원 함수이다. 비선형함수의 근사화에 대한 가장 간단한 접근 중에 하나는 일반적으로 고정된 비선형 기초함수(basis function) $\varphi_i(x)$의 선형결합에 의해 표현된다. 즉, 비선형함수의 근사화 식은

$$F(x) = \sum_{i=1}^{h} w_i \varphi_i(x) \tag{1}$$

이다. 이때 신경망의 학습목적은 계수집합인 연결가중치 w_i를 추정하는 것이다. 그러나 요구되는 기초함수의 개수는 함수의 차원에 따라 지수 함수적으로 증가된다. radial basis 함수는 문제의 차원이 증가함에 따라 조정되는 파라미터의 확산문제를 해결할 수
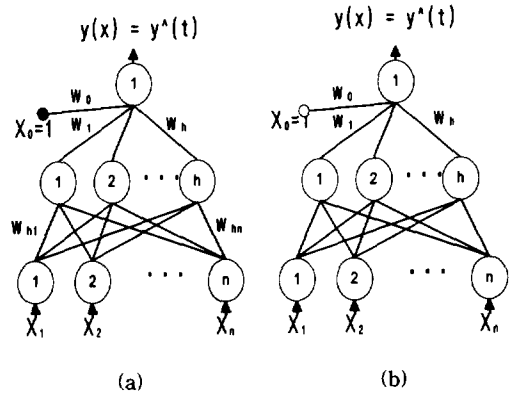
그림 1. (a) Multilayer perceptron 및 (b) radial basis 함수 3층전향 신경망

있는 다차원 근사화를 위한 강력한 방법을 제공한다. 일반적으로 이용되고 있는 전형적인 radial basis 함수 $\varphi(x) = \Phi(\| x - c \|)$로는 가우스함수(Gaussian function, $\Phi(r) = \exp(-\gamma^2/\sigma^2)$), thin plate splines ($\Phi(r) = \gamma^2 \log(\gamma)$),그리고 다중 2차 함수(multiquadratic function, $\Phi(r) = (\gamma^2 + \sigma^2)^{1/2}$) 등이 있다. 이들 식에서 σ는 거리의 계수(scaling) 파라미터인 평활요소이다. 일반적으로 radial basis 함수로는 중심 **c**에서 첨두치를 가지며 그 중심으로부터 거리가 증가할수록 단조 감소하는 값을 가지는 가우스함수가 가장 널리 이용되고 있다.

한편, RBF 신경망은 MLP와는 달리 하나의 은닉층만을 가지며, 은닉층의 연결은 가중되지 않는다. 즉 각각의 은닉층 뉴런은 각 입력값을 그대로 받아 radial basis 함수를 수행하는 처리장치이다. 일반적으로 radial basis 함수는 MLP의 단조 s-자형(monotonic sigmoid) 함수와 대조적으로 비단조 함수가 이용된다. 출력층의 뉴런은 MLP와 유사하게 입력들의 단순한 가중된 합만을 계산한다. 그림 1의 (b)에서 RBF 신경망의 출력층 뉴런의 출력은 radial basis 함수의 선형조합을 계산함으로써 얻어질 수 있다. 즉,

$$y^{\wedge}(t) = w_0 + \sum_{i=1}^{h} w_i \Phi_i(\| x(t) - c_i \|) \tag{2}$$

이다. 여기서 t는 반복수이고, $x(k) = [x_1, x_2, \ldots, x_n]^T$로 입력 데이터들이며, $c_i \in R^n$ ($i = 1, 2, .., h$)으로 함수의 중심이다. 만약 충분한 개수의 은닉층 뉴런들이 이용되고 중심들이 입력패턴의 영역에 적당하게 분

포되어 있다면, RBF 신경망은 다양한 부류의 비선형 다차원 함수들을 근사화할 수 있을 것이다. 그러나 RBF 신경망의 성능은 radial basis 함수의 중심, 평활요소, 그리고 연결가중치에 따라 변하지만 그 중에서도 선택된 함수의 중심값에 가장 민감하게 변한다.

radial basis 함수의 파라미터들은 3개의 연속적인 단계로 결정될 수 있다. 첫 단계로, h개의 중심벡터 c_i가 결정되어야 한다. 이를 위한 가장 간단한 방법으로는 학습 데이터의 부분집합으로부터 랜덤하게 선택하는 것이다. 그러나 이 방법에서는 전체 입력영역을 다 포함시키키기 위해서는 상대적으로 많은 수의 은닉층 뉴런이 필요하게 되며 이들의 중심값을 계산하기 위한 계산량도 더불어 증가하게 된다. 또 다른 널리 이용되고 있는 방법으로는 k-평균 군집화 알고리즘이 있다. 이 알고리즘의 기본적인 생각은 입력데이터의 밀도가 높아지면 중심의 밀도도 함께 높아진다는 것이다. 이 알고리즘은 군집의 중심집합과 학습 데이터의 분할을 찾는 병렬형 알고리즘이다. 각 군집의 중심은 RBF 신경망에서 h개의 은닉층 뉴런 중에 하나와 관계된다. 즉, 데이터는 그것과 가장 가까운 중심을 가진 군집에 할당되는 방법으로 분할된다. 그러나 이 알고리즘 역시 초기의 군집 중심을 어떻게 설정하는가에 따라 그 성능이 크게 변한다. 다음 단계로는 구해진 radial basis 함수의 중심벡터를 이용하여 평활요소 σ_i를 구하는 것으로 이는 P-최근접 이웃(P-nearest neighbour)으로부터 결정될 수 있다. 즉,

$$\sigma_i = \frac{1}{P}\sum_{j=1}^{P}[\, \| c_i - c_j \|^2]^{1/2} \tag{3}$$

이다. 여기서 c_j는 σ_i의 P-최근접 이웃이다. 본 연구에서는 σ_i의 값에 따른 성능을 분석하기 위해서 임의 상수값으로 설정하여 실험하였다. 마지막 단계는 설정된 radial basis 함수의 중심벡터 c_i와 평활요소 σ_i을 이용하여 식 (1)과 같은 선형회귀(linear regression)를 해결함으로써 연결가중치 w_i을 갱신하는 것이다. 한 예로 선형최소자승의 표준을 적용할 때 간단한 최소평균제곱(least mean square : LMS) 알고리즘을 얻을 수 있다. 즉, 연결가중치 w_i의 갱신은

$$\frac{dw_i}{dt} = \eta_i\, e_t\, \varphi_i[x(t)] = \eta_i[y(t) - (\,w_o + \sum_{i=1}^{h} w_i \Phi(\| x(t) - c_i \|)]\Phi(\| x(t) - c_i \|) \tag{4}$$

을 이용하여 이루어 진다. 여기서 η_i는 학습율로서 양수값이며, $e_k = y(k) - y^{\wedge}(k)$이다.

따라서 근본적으로 입력 데이터가 가지는 차원을 감소시킬 수 있는 새로운 기법을 도입함으로써 결정해야 될 은닉층의 뉴런 수 및 그 중심값과 같은 파라미터들의 설정이 용이하게 될 것이다. 즉, 주어진 입력 데이터의 차원을 감소시켜 그에 해당되는 특징값을 RBF 신경망의 은닉층 파라미터로 사용함으로써 이들을 결정하는데 요구되는 계산량을 효과적으로 감소시킬 수 있을 것이다.

2.2 적응적 특징추출을 위한 PCA

PCA는 주어진 입력 데이터의 특징을 추출하는 기법으로 널리 이용되고 있으며, 이는 정상확률과정(stationary stochastic process)의 입력벡터와 특별한 차원의 출력벡터를 위한 최적의 선형변환(linear transform)을 결정하는 통계적 방법이다[4-7]. 결국 입력 데이터 내에 포함된 정보를 추출하고 압축하여 통계적 규칙들을 찾아내는 것이다. 따라서 PCA는 대용량의 입력데이터를 통계적으로 독립인 특징들의 집합으로 변환시키는 기법이다.

PCA는 n 차원 입력공간의 데이터를 m 차원 출력공간의 데이터로 투영시키는 것이다. 여기서 $m \ll n$이며, 이는 입력데이터 벡터의 대부분 내부정보를 유지하도록 차원의 감소를 얻는 것이다. 자기상관행렬(auto-correlation matrix) $\mathbf{R_{xx}} = \langle \mathbf{xx}^T \rangle$를 가진 평균이 영인 입력벡터 $\mathbf{x} = [\, x_1, x_2, \cdots, x_n\,]^T$에 대해서 생각해 보자. 여기서 $\langle \mathbf{x} \rangle$는 기대치를 나타낸다. 또한 $\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_m$ 이 $\mathbf{R_{xx}}$ 의 고유벡터(eigenvector)와 직교되는 연결가중치 벡터라 할 때, $\mathbf{w}_1 = [\, w_{11}, w_{12}, \cdots, w_{1n}\,]^T$는 가장 큰 고유치(eigenvalue) λ_1과 일치하며, $\mathbf{w}_2 = [\, w_{21}, w_{22}, \cdots, w_{2n}\,]^T$는 두 번째로 큰 고유치 λ_2와 $\mathbf{w}_n = [\, w_{n1}, w_{n2}, \cdots, w_{nn}\,]^T$는 가장 작은 고유치 λ_n과 일치한다. 이상의 관계를 행렬방정식으로 나타내면 다음과 같다. 즉,

$$\mathbf{R}_{xx}\, \mathbf{w}_j = \lambda_j\, \mathbf{w}_j \quad (j = 1, 2, \cdots, n) \tag{5}$$

이다. 여기서 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ 이다.

주어진 입력벡터 $\mathbf{x}$를 위한 첫 번째 m 개의 주요 특징벡터 $\mathbf{y}$는 다음의 선형변환식으로 나타낼 수 있다. 즉,

$$\mathbf{y} = \mathbf{Wx} \tag{6}$$

이다. 여기서 $\mathbf{W} = [\, \mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_m \,]^T \in \mathbb{R}^{m \times n}$ 이다. 이 식은 행렬 $\mathbf{W}$의 행이 가장 큰 고유치와 일치하는 상관행렬 $\mathbf{R}_{xx}$의 고유벡터임을 의미한다. 다시 말하면, 입력 데이터 공간에서 m 차원의 주요특징을 나타내는 부공간은 $\mathbf{R}_{xx}$의 m 개 주요 고유벡터에 의해 구성된 부공간으로 정의된다. 결국 PCA는 $< \| \mathbf{w}_j^T \mathbf{x} \|_2^2 >$가 최대인 고유벡터 $\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_m$ 의 방향을 찾는 것이다.

한편, 입력되는 데이터의 주요특징들을 추출하기 위한 기존의 방법들을 살펴보면, 수치적 방법에서는 먼저 자기상관행렬 $\mathbf{R}_{xx}$를 계산한 다음 그것과 일치하는 고유벡터를 계산한다. 그러나 입력데이터 벡터가 대규모의 차원을 가진다면, 상관행렬은 대단히 크게 되며 또한 요구되는 고유벡터를 찾는 것도 매우 복잡한 일이다. 결국 기존의 수치적 기법들은 대규모의 실시간 처리가 요구되는 응용문제에서는 매우 비효율적이다. 이를 해결하기 위한 대안으로 상관행렬의 고유벡터를 실시간으로 추정하기 위하여 적응학습 알고리즘의 신경망을 이용하는 방법들이 제안되었다. 여기서는 입력 데이터 벡터로부터 직접 고유벡터를 추정할 수 있기 때문에 상관행렬의 추정과정이 요구되지 않는다. 이때 이용된 신경망은 주로 입력층과 출력층으로 구성된 단층구조이다. Oja[4]는 정규화된 헤비안규칙(normalized Hebbian rule)의 적응학습 방법을 이용한 간단한 뉴런모델을 제안하여 정상입력벡터 계열의 주요 특징으로 연결가중치 벡터(weights vector)가 수렴됨을 증명하였다. 그러나 이 방법에서는 한 개의 중요한 주요특징만을 추출할 수 있으며, 임의의 m 개의 특징추출로의 확장은 불가능하였다. Sanger 등[4]은 일반화된(generalized) 헤비안규칙을 이용함으로써 정상과정의 m 개의 가장 중요한 주요특징들을 계산하기 위한 다중 뉴런모델을 제안하였다. 한편, Foldiak[4,5]은 망의 입력과 출력사이의 연결가중치 경신에는 정규화된 헤비안규칙을 이용하고, 망의 출력사이의 측면연결 가중치 경신에는 반 헤비안규칙(anti-Hebbian rule)을 함께 이용한 학습알고리즘을 제안하였다. 최근 Bannour 등[6]은 Gram-Schmidt 직교화와 순환최소자승(recursive least square)법을 이용한 학습알고리즘을 제안하였다. 여기서는 단층신경망을 대상으로 Sanger에 의해 제안된 학습 알고리즘과 Foldiak에 의해 제안된 학습 알고리즘에 대해서 알아본다.

2.3 Sanger과 Foldiak의 학습알고리즘에 의한 적응적 특징추출

주어진 입력벡터에 대한 첫 번째로 주요한 특징을 추출하기 위해서는 출력장치로 한 개의 선형 처리장치를 가진 망이 이용된다. 이를 위한 많은 학습알고리즘들이 제안되었으며, 이때의 학습은 자율학습이다. m 개의 주요특징을 추출하기 위해서는 m 개의 선형 처리장치로 구성되는 망이 요구되나 한 개의 선형 처리장치를 가진 망을 위한 학습알고리즘을 확장하여 여러 개의 선형 처리장치를 가진 망을 학습시키는 것은 불가능하다. 따라서 주어진 입력벡터에 대한 m 개의 첫 번째 주요특징을 추출하기 위해서는 m 개의 선형 처리장치를 출력층으로 하는 단층신경망이 구성되어야 하며, 이 망을 위한 별도의 학습알고리즘이 요구된다. 그림 2는 m 개의 첫 번째 주요특징 추출을 위해 n 개의 입력뉴런과 m 개의 출력뉴런으로 구성된 단층신경망의 구조이다. 그림에서 입력과 출력의 관계를 행렬방정식으로 나타내면 식 (6)과 동일하다. 여기서 각 뉴런의 활성화함수는 선형함수이고, $\mathbf{W} = [\, \mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_m \,]^T$ 로 연결가중치 행렬이며, $\mathbf{w}_i = [\, w_{i1}, w_{i2}, \cdots, w_{in} \,]^T$는 연결가중치 벡터이다. 따라서 신경망을 학습시켜 연결가중치를 구함으로써 주어진 n 개의 입력벡터 $\mathbf{x}$로부터 m 개의 주요특징벡터 즉 출력벡터 $\mathbf{y}$를 구할 수 있다.

일반화된 헤비안규칙에 기반을 둔 그림 2의 단층신경망을 학습시키는 알고리즘은 다음과 같다. 즉,

$$w_{ij}(t+1) = w_{ij}(t) + \eta_i \, y_i(t) \, [\, x_j(t) - \sum_{h=1}^{i} w_{hj}(t) \, y_h(t) \,], \; (\, i = 1,2,\cdots,m, \; j = 1,2,\cdots,n) \qquad (7)$$

이다. 여기서 η_i는 학습율로서 학습 반복수 t가 증가함에 따라 보통 지수 함수적으로 감소된다. 식 (7)을 이용하여 연결가중치를 구한 다음 입력벡터와 함께 식 (6)에 대입하면 주요특징 벡터를 구할 수 있다. 즉, 학습은 상관행렬의 고유벡터를 추정하기 위하여 임의의 초기연결 가중치로부터 주요특징을 내포하는 연결 가중치를 구하는 과정이다. 결국 신경망은 입력데이터 내에 포함된 상호독립인 정보를 부호화시키는 검출기이다.

한편, 입력뉴런과 출력뉴런 사이의 연결 및 출력뉴런 상호간에도 연결을 가진 새로운 단층신경망이 m 개의 첫 번째 주요특징을 추출하기 위해 제안되었
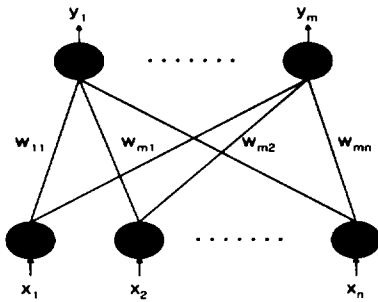
그림 2. m 개의 첫 번째 주요특징 추출을 위한 단층신경망

다[2]. 그림 3은 n 개의 입력뉴런과 m 개의 출력뉴런으로 구성된 입력과 출력뉴런간 및 출력뉴런 상호간의 측면연결을 가진 단층신경망의 구조이다. 그림에서 입력과 출력의 관계를 나타내면 다음과 같다. 즉,

$$y_i = \sum_{j=1}^{n} w_{ij}\, x_j + \sum_{h=1}^{i-1} u_{ih}\, y_h \quad (i = 1,2,\cdots,m) \qquad (8)$$

이다. 여기서 w_{ij}는 입력뉴런과 출력뉴런을 연결하는 연결가중치이고, u_{ih}는 출력뉴런 상호간의 측면연결가중치이다. 이때 출력뉴런 간의 상호연결을 보면 뉴런 i는 다만 h<i인 뉴런에만 연결된다. 따라서 신경망을 학습시켜 입출력 뉴런간 및 출력뉴런 상호간의 연결가중치를 각각 구함으로써 주어진 n 개의 입력벡터 **x**로부터 m 개의 주요특징벡터 즉 출력벡터 **y**를 구할 수 있다.
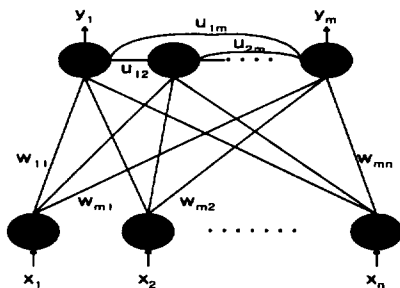
그림 3. m 개의 첫 번째 주요특징 추출을 위한 측면연결의 단층신경망

그림 3과 같은 단층신경망을 학습시키기 위해서 정규화된 헤비안규칙과 반 헤비안규칙에 기반을 둔 알고리즘이 이용된다. 입력과 출력뉴런 간의 연결가중치 경신에는 정규화된 헤비안규칙을 이용하고 출력뉴런 간의 측면 연결가중치 경신에는 반 헤비안규칙을 이용한다. 먼저 입출력 뉴런간의 연결가중치 w_{ij}의 경신규칙을 살펴보면 다음과 같다. 즉,

$$w_{ij}(t+1) = w_{ij}(t) + \eta\, [\, y_i(t)\, x_j(t) - w_{ij}(t)\, y_i(t)^2 \,], \quad (i = 1,2,\cdots,m,\ j = 1,2,\cdots,n) \qquad (9)$$

이다. 또한 출력뉴런 간의 측면 연결가중치 u_{ih}의 경신규칙은 다음과 같다. 즉,

$$u_{ih}(t+1) = u_{ih}(t) + \rho\, y_i(t)\, y_h(t), \quad (i > h) \qquad (10)$$

이다. 여기서 ρ는 식 (9)의 η와 일반적으로 동일한 값이다. 따라서 식 (9)와 (10)을 이용하여 입력과 출력뉴런 간의 연결가중치 및 출력뉴런 상호간의 측면 연결가중치를 각각 경신시켜 식 (8)에 대입하면 m 개의 주요특징들을 추출할 수 있다.

결국 PCA는 높은 차원의 입력공간을 더 낮은 차원의 표현공간으로 사상시켜 입력데이터가 가지는 두드러진 특징들을 추출하는 기법이다. 이는 입력 데이터 내에 존재하는 주요특징들을 추출함으로써 입력의 개수를 감소시키는데 이용될 수 있다. 따라서 RBF 신경망의 은닉층 뉴런 개수는 원래의 입력데이터를 대상으로 PCA에 의해 얻어지는 감소된 주요특징의 개수로 설정하고, 은닉층 뉴런의 중심값은 얻어진 주요 특징값으로 설정할 수 있을 것이다. 이렇게 한다면 RBF 신경망에서의 은닉층 파라미터인 뉴런의 개수와 함수의 중심값 설정이 용이하게 되어 이 신경망이 가지는 우수한 속성을 그대로 살릴 수 있을 것이다. 일반적으로 Sanger 등에 의해 제안된 학습규칙보다는 Foldiak 에 의해 제안된 학습규칙이 수렴속도면에서 더 우수한 것으로 알려 있다.

3. 시뮬레이션 결과 및 분석

제안된 PCA 기법을 이용한 RBF 신경망의 성능을 평가하기 위해서 3층 전향 신경망을 구성하였다. RBF 신경망에서의 은닉뉴런과 출력뉴런 사이의 초기 연결가중치와 PCA를 위한 단층신경망의 입력뉴런과 출력뉴런 및 출력 뉴런사이의 측면연결을 위한 초기 연결가중치는 각각 랜덤시드(random seed)를 이용하여 -1에서 +1 사이의 임의 값으로 설정하였다. 학습은 전체 반복회수가 20,000이상이거나 전체 오차가 허용치 이하일 때, 또는 전체 오차함수 값의 변화가 10^{-6}이하일 때 종료되도록 하였다.

제안된 RBF 신경망에서 PCA 기법으로는 Foldiak의 방법을 이용하였다. PCA 기법을 이용한 RBF 신경망을 실제 암환자를 2 부류(초기와 악성)로 분류하는 문제를 대상으로 펜티엄Ⅱ-MMX200 (㎒) 컴퓨터를 이용하여 시뮬레이션한 후 그 타당성을 확인하였으며, 지금까지 널리 이용되고 있는 k-평균 군집화 알고리즘을 이용한 RBF 신경망에 의한 결과와 비교 고찰하였다.

실험에 이용한 암환자 분류는 많은 연구 논문에서의 벤치마크 문제로 이용되었으며[2-4], 그 데이터는 미국 위스콘신 대학병원에서 제공되는 "Wisconsin breast cancer databases"[10]을 대상으로 하였다. 원 데이터베이스는 699명의 환자를 대상으로 9 개의 정수 값(1 - 10)으로 표현된 조건속성과 초기와 말기(초기 : 2, 말기 : 4)를 나타내는 1 개의 분류속성으로 구성된다. 그리고 699명중에서 16명의 환자에 대해서는 그 속성 값들을 빠뜨렸다. 본 실험에서는 속성 값이 나타나 있지 않은 16명의 환자를 제외한 나머지 683명의 환자 중 임의의 200명을 대상으로 하고, 100명은 학습을 위한 데이터로 나머지 100명은 시험을 위한 데이터로 이용하였다. 또한 실험에 이용된 데이터들에서 환자의 식별번호인 첫 번째 속성은 제거하였고, 모든 조건속성들은 정수 10으로 나누었으며, 분류속성은 2(초기)이면 0으로 4(말기)이면 1로 표현하였다. 학습이 종료된 후 출력뉴런의 값이 0.1 이하이면 0(초기)으로 하고 0.9 이상이면 1(말기)로 판정하였다. 다음의 그림 4는 5명의 환자를 대상으로 실제의 데이터와 본 실험에서 사용되는 데이터를 나타내는 예이다.

1081791, 6, 2, 1, 1, 1, 1, 7, 1, 1, 2
1084584, 5, 4, 4, 9, 2, 10, 5, 6, 1, 4
1091262, 2, 5, 3, 3, 6, 7, 7, 5, 1, 4
1096800, 6, 6, 6, 9, 6, ?, 7, 8, 1, 2
1099510, 10, 4, 3, 1, 3, 3, 6, 5, 2, 4

(a) 실데이터

0.6, 0.2, 0.1, 0.1, 0.1, 0.1, 0.7, 0.1, 0.1, 0
0.5, 0.4, 0.4, 0.9, 0.2, 1, 0.5, 0.6, 0.1, 1
0.2, 0.5, 0.3, 0.3, 0.6, 0.7, 0.7, 0.5, 0.1, 1
0.6, 0.6, 0.6, 0.9, 0.6, ?, 0.7, 0.8, 0.1, 0 (실험에서 제외)
1, 0.4, 0.3, 0.1, 0.3, 0.3, 0.6, 0.5, 0.2, 1

(b) 실험용 데이터

그림 4. 5명의 환자에 대한 속성데이터 예

그림에서 ?는 그 속성 데이터가 알려져 있지 않아 실험에서는 학습 데이터로 이용되지 않은 경우이다.

그림 5는 평활요소 σ(σ=1.0와 신경망의 초기 연결가중치(랜덤시드 = 0)를 고정하고 은닉층 뉴런수의 변화와 k-평균 군집화 및 제안된 알고리즘 각각에서의 수렴률과의 관계를 나타낸 것이다. 여기서 평활요소 값과 랜덤시드 값의 설정은 실험을 통하여 얻어진 값들이다. 그림에서 보면, 제안된 알고리즘을 이용한 신경망의 수렴특성이 k-평균 군집화 알고리즘을 이용한 신경망에 비해 다소 우수한 수렴특성이 있음을 보여준다. 이는 주어진 문제에 대해서 제안된 알고리즘이 k-평균 군집화 알고리즘에 비하여 신경망의 은닉층 뉴런의 중심값을 더욱 잘 설정한 것으로 추측된다. 또한 두 알고리즘 모두의 경우에 은닉층 뉴런의 수가 증가할수록 그 수렴률도 함께 증가됨을 알 수 있다. 그러나 은닉층 뉴런수의 증가는 계산량의 증가를 가져오며 결국 신경망을 학습시간도 오래 걸리게 된다. 따라서 은닉층 뉴런의 수는 적당하게 설정되어야 할 것이다. 본 실험에서는 은닉층 뉴런의 수를 18 개로 할 때에 학습시간과 그 수렴률 측면에서 가장 효과적임을 확인할 수 있었다.

표 1은 은닉층 뉴런수(18 개)와 신경망의 초기 연결가중치(랜덤시드 = 0)를 고정하고 평활요소의 변화에 따른 두 알고리즘의 수렴성능을 비교한 것이다. 표에서 N_k와 N_p는 각각 k-평균 군집화와 제안된 알고리즘에 의한 신경망의 학습회수를 나타내며, t_k와 t_p는 신경망의 학습시간으로 CPU 시간을 나타낸 것이다. 또한 P_r은 학습 후에 100 개의 시험 데이터로 시험할 경우 올바르게 분류된 비율이며, E_i는 학습종료시의 평균자승 오차값을 나타낸 것이다. 표에서
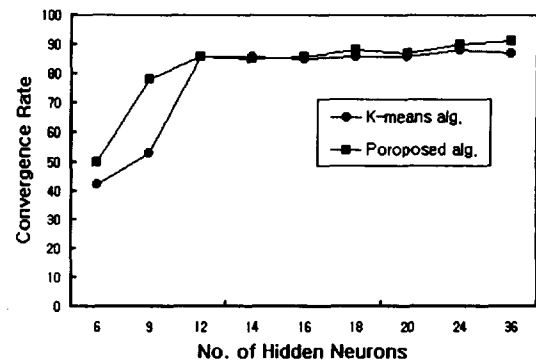
그림 5. 은닉층 뉴런수의 변화에 따른 수렴률

표 1. 평활요소의 변화에 따른 수렴성능 비교

σ_i	K-means alg.			Proposed alg.		
	$N_k(t_k)$	$P_r(\%)$	E_k	$N_p(t_p)$	$P_r(\%)$	E_k
0.01	2(0)	50	25.000000	0(0)	50	25.000000
0.1	839(14)	50	24.995972	0(0)	50	25.000000
0.5	20000(296)	68	4.160935	6396(98)	72	3.871657
1.0	5640(89)	89	0.344699	1499(23)	88	0.336547
2.0	3(1)	50	78110.164	304(4)	100	0.022349
2.5	0(0)	50	114369725.00	20000(296)	49	18145.125

N_r : no. of iterations, t_r : CPU time(sec), p_r : convergence rate(%), E_r : mean square error

보면, 제안된 알고리즘이나 k-평균 군집화 알고리즘을 이용한 신경망의 수렴성능은 평활요소 σ_i에 매우 의존함을 알 수 있다. 이는 Gaussian radial basis 함수의 폭을 결정하는 파라미터로 신경망의 정확도에 매우 많은 영향을 미침을 알 수 있다. 또한 표에서 제안된 알고리즘이 k-평균 군집화 알고리즘보다 우수한 수렴특성이 있음을 알 수 있으며, 평활요소의 변화범위도 더 넓음을 확인할 수 있다. 즉, 제안된 알고리즘의 경우는 평활요소의 변화가 0.5에서 2.0 사이의 범위를 가지나 k-평균 군집화 알고리즘의 경우는 0.5에서 1.0 사이의 범위를 가진다. 이는 제안된 알고리즘이 상대적으로 k-평균 군집화 알고리즘에 비해 평활요소에 덜 민감함을 알 수 있다. 특히, 실험에서는 평활요소 값을 2.0으로 하고 은닉층 뉴런수를 18 개로 하여 신경망의 초기 연결가중치를 100번 바꾸어 가면서 학습시킨 결과, 평균적으로 수렴회수는 655.27, 수렴시간은 9.82, 그리고 시험 데이터 100개에 대해 모두 정확하게 분류하는 결과를 얻을 수 있었다. 이는 k-평균 군집화 알고리즘에 비해 상대적으로 PCA를 이용하는 제안된 알고리즘의 우수성을 보여주는 것이다. 한편, 표의 결과로부터 평활요소의 값에 따라 평균자승 오차값 E_i가 각각 24.995972, 25.000000, 그리고 매우 큰 값을 가지는 이러한 경우는 학습이 이루어지지 않는 발진하는 결과를 보여준 것이며, 이 때의 수렴률은 두 부류(0 또는 1) 중 어느 한 부류로 항상 고정된 경우이다.

표 2는 초기 연결가중치를 100번 변화시켜가면서 RBF 신경망을 학습시킨 결과, 학습에 소요된 반복회수 N_i, CPU 시간 t_i, 그리고 평균자승 오차값 E_i 의 평균 $\bar{x}$와 표준편차 μ를 각각 나타낸 것이다. 또한 은닉층 뉴런의 개수와 평활요소는 두 알고리즘 모두에 대해 각각 18 개와 1.0으로 설정하였으며, 각 시도는 랜덤시드를 변화시켜 수행하였다. 여기서도 P_r은 학습 후에 100 개의 시험 데이터로 시험할 경우 올바르게 분류된 비율을 나타낸 것이다. 표에서 보는 것처럼, 제안된 알고리즘을 이용한 RBF 신경망이 k-평균 군집화 알고리즘을 이용한 신경망보다 학습시간과 데이터의 분류률 면에서 더욱 우수한 성능이 있음을 알 수 있다. 특히 분류률에서는 제안된 알고리즘이 약간 우수한 성능을 보이나 학습시간 면에서는 약 3배정도 빠름을 알 수 있다. 표에서 보면 두 알고리즘을 이용하는 신경망 모두 뉴런의 초기 연결가중치에 따라 그 성능이 달라짐을 알 수 있다. 이는 이용된 RBF 신경망의 학습이란 은닉층 뉴런과 출력층 뉴런을 연결하는 출력층 연결가중치를 적응적으로 변화시키는 것으로 여기서는 기울기하강법을 이용하였다. 결국 이로 인해 학습율의 변화에 따라 학습성능도 달라진다. 그러나 가장 최적의 학습율을 설정할 수 있는 방법은 알려져 있지 않으며 본 실험에서는 여러 가지 학습율 값으로 실험 후에 학습율을

표 2. 100개의 학습 데이터를 대상으로 100번 시도에 따른 실험 결과

	K-means alg.			Proposed alg.		
	N_k	t_k	E_k	N_p	t_p	E_p
$\bar{x}$	6194.69	97.27	0.368058	1966.80	30.64	0.335882
μ	786.75	26.11	0.045847	173.12	2.81	0.000326
P_r	86			88		

$\bar{x}$: means, μ : standard deviation, N_r : no. of iterations, t_r : CPU time(sec), p_r : convergence rate(%), E_r : mean square error

0.1/(log_{10}(N))로 설정하였다. 이는 학습 초기에는 큰 값을 가지나 학습 반복수가 증가함에 따라 감소되는 값으로 그 수렴속도를 증가시키기 위함이다. 한편, 표준편차의 경우에 제안된 알고리즘의 값이 k-평균 군집화 알고리즘의 값보다 상대적으로 더 작은 값을 가져 제안된 알고리즘이 초기 연결가중치에 대한 의존도도 적음을 추정할 수 있다.

4. 결 론

본 논문에서는 PCA 기법을 RBF 신경망에 이용하는 새로운 조합형 기법을 제안하였다. 여기서 PCA 기법은 주어진 입력 데이터가 가지는 차원을 감소시켜 그에 해당하는 특징값을 얻을 수 있는 효율적인 기법이며, 얻어진 특징값으로 RBF 신경망의 은닉층 파라미터를 결정하는 것이다. 이렇게 하면 PCA 기법이 가지는 대용량의 입력데이터를 통계적으로 독립인 특징들의 집합으로 변환시키는 장점과 RBF 신경망이 가지는 우수한 속성을 그대로 살릴 수 있다.

제안된 기법의 RBF 신경망을 200명의 암환자를 2부류(초기와 악성)로 분류하는 문제에 적용하여 시뮬레이션한 결과, k-평균 군집화 알고리즘을 이용한 RBF 신경망에 의한 결과와 각각 비교할 때 학습시간과 시험 데이터의 분류률 면에서 더욱 우수한 성능이 있음을 확인할 수 있었다. 특히, 초기 연결가중치의 설정에 대한 융통성과 은닉층 뉴런의 평활요소 설정 여유도 측면에서도 더욱 우수함을 알 수 있었다.

향후 제안된 알고리즘의 RBF 신경망을 좀 큰 규모의 문제와 통신이나 영상 인식 등과 같은 좀 더 다양한 분야에의 응용에 대한 연구가 계속 진행되어야 할 것이며, RBF 신경망의 병렬구현에 대한 연구도 남아 있다.

참 고 문 헌

[1] J. T. Tou and R. C. Gonzalez, 'Pattern Recognition Principles,' Addison-Wesley Pub., London, 1974

[2] P. D. Wasserman, 'Advanced Methods in Neural Computing,' Van Nostrand Reinhild, New York, 1993

[3] S. Haykin, 'Neural Networks : A Comprehensive Foundation,' IEEE Press, New York, 1994

[4] A. Cichock and R. Unbehaun, 'Neural Networks for Optimization and Signal Processing,' John Wiley & Sons., New York, 1993

[5] P. Foldiak, "Adaptive Network for Optmal Linear Feature Extraction," International Joint Conference on Neural Networks, Washington D. C., vol 1, pp. 401-406, June 1989

[6] S. Bannour, A. Mahmood, and A. Sadjadi, "Principal Component Extraction Using Recursive Least Squares Learning," IEEE Trans. on Neural Networks, vol. 6, no. 2, pp. 457-469, March 1995

[7] K. I. Diamantaras and S. Y. Kung, 'Principal Component Neural Networks : Theory and Applications, Adaptive and Learning Systems for Signal Processing, Communications, and Control', John Wiley & Sons, Inc., 1996

[8] J. Park and I. Sandberg, "Universal Approximation Using Radial Basis Function Networks," Neural Computatio, vol 3, no. 2, pp. 246-257, 1989

[9] A. Baradi and N. A. Borghese, "Learning from Data : General Issuses and Special Applications of Radial Basis Function Networks," Technical Report TR-98-028, Berkeley, 1998

[10] W. H. Wolberg, "Wisconsin Breast Cancer Database," Univ. of Wisconsin Hospital, July 1992(ftp://mlftp.diee.unica.it/pub/mac...sin/breast-cancer-wisconsin.data)

조 용 현

1979년 2월 경북대학교 공과대학 전자공학과(공학사)
1981년 2월 경북대학교 대학원 전자공학과(공학석사)
1993년 2월 경북대학교 대학원 전자공학과(공학박사)
1983년 9월~1984년 2월 삼성전자(주) 사원
1984년 3월~1987년 2월 한국전자통신연구소 연구원
1987년 3월~1997년 2월 : 영남전문대학 전자과 부교수
1997년 3월~현재 대구효성가톨릭대학교 컴퓨터정보통신공학부 조교수
관심분야 : 신경망, 병렬분산처리, 신호처리, 영상인식, 교환기 등
E-mail : yhcho@cuth.cataegu.ac.kr